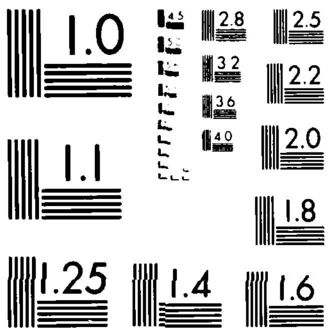
AD-A152 317 PERFORMANCE EVALUATION TESTS FOR ENVIRONMENTAL RESEARCH (PETER) EVALUATION OF 112 MEASURES(U) NAVAL BIODYNAMICS LAB NEW ORLEANS LA A C BITTNER ET AL. SEP 84 1/1

UNCLASSIFIED NBDL-84R006 F/G 5/10 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

NBDL-84R006

PERFORMANCE EVALUATION TESTS FOR ENVIRONMENTAL RESEARCH (PETER): EVALUATION OF 112 MEASURES

Alvah C. Bittner, Jr., Robert C. Carter, Robert S. Kennedy
Mary M. Harbeson, and Michele Krause

September 1984

NAVAL BIODYNAMICS LABORATORY
New Orleans, Louisiana

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NBDL-84-R006	2. GOVT ACCESSION NO. AD-A152317	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Performance Evaluation Tests for Environmental Research (PETER): Evaluation of 112 Measures		5. TYPE OF REPORT & PERIOD COVERED Research Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A. C. Bittner, Jr., R. C. Carter, R. S. Kennedy, M. M. Harbeson, and M. Krause		8. CONTRACT OR GRANT NUMBER(s) Naval Training Equipment Center; Task No 3775-2P4
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Biodynamics Laboratory P. O. Box 29407 New Orleans, LA 70189-0407		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command Naval Medical Command, National Capital Region Bethesda, MD 20814		12. REPORT DATE September 1984 13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Behavioral Science, Psychology, Human Factors Engineering, Human Performance, Repeated Measures, Psychological Tests, Cognitive Tests, Psychomotor Tests, Perceptual Tests, Test Construction, Environmental Tests, Test Batteries.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See back of form.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The goal of the Performance Evaluation Tests for Environmental Research (PETER) Program was to identify a set of measures of human cognitive, perceptual, and motor capabilities for use in the study of environmental and other time-course effects. Tasks were evaluated as suitable for repeated measures applications when their intertrial means, variances, and correlations were well-behaved under constant baseline conditions. The results of this program are documented in more than 90 reports. Because of the volume of this literature, a review is needed to enhance the applicability of the results. This report provides an evaluation of 112 measures studied in the PETER Program. They are categorized into four groups based upon consideration of task stability and task definition. The Recommended category contained 30 measures that clearly obtained total stabilization and had an acceptable level of reliability efficiency (i.e., $r_{xx} \geq .50$, normalized to a three minute administration). The Acceptable-But-Redundant category contained 15 measures that met the same requirements as those in the Recommended category but were found redundant. The 35 measures in the Marginal category usually had desirable features which were outweighed by faults. The 32 measures in the Unacceptable category were characterized by either differential instability or weak reliability efficiency ($r_{xx} < .15$). This category contained an inordinate number of slope and other derived measures. Characteristics of the measures are presented in application oriented tables. Measures suitable (or unsuitable) for repeated measurements are identified and compared. It is our opinion that the 30 measures in the Recommended category should be given first consideration for environmental research applications. Further, it is recommended that information pertaining to preexperimental practice requirements and stabilized reliabilities should be utilized in repeated measures environmental studies.

NBDL-84R006

PERFORMANCE EVALUATION TESTS FOR ENVIRONMENTAL RESEARCH (PETER): EVALUATION OF 112 MEASURES

Alvah C. Bittner, Jr., Robert C. Carter, Robert S. Kennedy
Mary M. Harbeson, and Michele Krause

September 1984

Naval Training Equipment Center
Task No. 3775-2P4

Approved by	Released by
J. C. Guignard Chairman, Editorial Review Board	Captain R. J. Biersner, MSC, USN Commanding Officer

Naval Biodynamics Laboratory
Box 29407
New Orleans, LA 70189-0407

SUMMARY PAGE

PROBLEM

The goal of the Performance Evaluation Tests for Environmental Research (PETER) Program was to identify a set of measures of human cognitive, perceptual, and motor capabilities for use in the study of environmental and other time-course effects. Tasks were evaluated as suitable for repeated measures applications when their intertrial means, variances, and correlations were well-behaved under constant baseline conditions. The results of this program are documented in more than 90 reports. Because of the volume of this literature, a review is needed to enhance the applicability of the results.

FINDINGS

This report provides an evaluation of 112 measures studied in the PETER Program. They are categorized into four groups based upon consideration of task stability and task definition. The Recommended category contained 30 measures that clearly obtained total stabilization and had an acceptable level of reliability efficiency (i.e., $r_{xx} \geq .50$, when normalized to a three minute administration). The Acceptable-But-Redundant category contained 15 measures that met the same requirements as the Recommended, but were found redundant. The 35 measures in the Marginal category usually had desirable features which were outweighed by faults. The 32 measures in the Unacceptable category were characterized by either differential instability or weak reliability efficiency ($r_{xx} < .15$). This category contained an inordinate number of slope and other derived measures. Characteristics of the measures are presented in application oriented tables. Measures suitable (or unsuitable) for repeated measurements are identified and compared.

RECOMMENDATIONS

It is our opinion that the 30 measures in the Recommended category should be given first consideration for environmental research applications. Further, it is recommended that information pertaining to preexperimental practice requirements and stabilized reliabilities should be utilized in repeated measures environmental studies.

ACKNOWLEDGEMENTS

Portions of this report were prepared under Naval Training Equipment Center Task No. 3775-2P4. Contributions by Dr. Kennedy were made under NTEC Contract No. N61339-81-C-0105 and NASA Contract No. NAS 9-16982. An abbreviated version of this report, entitled "Performance Evaluation Tests for Environmental Research (PETER): Good, Bad, and Ugly," was presented at the 28th Annual Meeting of the Human Factors Society in San Antonio, TX, in October 1984. LCDR Carter is now with Chief of Naval Operations, Washington, DC; Dr. Kennedy is with Essex Corp., Orlando, FL; and Ms. Krause is self-employed in Vero Beach, FL.

Trade names of materials or products of commercial or nongovernment organizations are cited where essential for precision in describing research procedures or evaluation of results. Their use does not constitute official endorsement or approval of the use of such commercial hardware or software.

PERFORMANCE EVALUATION TESTS FOR ENVIRONMENTAL RESEARCH (PETER):

EVALUATION OF 112 MEASURES

Performance Evaluation Tests for Environmental Research (PETER), a program to evaluate the suitability of human performance tests for repeated measures applications, has been underway since 1977 (50, 62). The goal of this program was to identify a set of measures of human cognitive, perceptual, and motor capabilities for use in the study of environmental and other, time-course effects. Environmental stressors, for example, those experienced in Navy workplaces such as aboard ship, may reduce well-being and productivity. The gross effects of such arduous environments are readily observable, but in order to detect more subtle effects, a sensitive measuring device is necessary. The PETER Battery has been designed to be sensitive to changes in performance and for other repeated measures applications.

Prior to the advent of the PETER Program, concerted efforts at research on the differential effects of practice on test characteristics had not appeared with any regularity in the recent literature (37, 59, 62). Yet it is only with such a paradigm that subtle changes in performance can be most efficiently detected (110). In previous battery development, attention was paid to the stability of the means, and to a lesser extent to the stability of the standard deviations or variances. The PETER Program focused also on the stability and reliability of the intertrial correlations (62).

Tasks were evaluated as suitable for inclusion in the battery when their intertrial means, variances and correlations were well-behaved under constant baseline conditions (62). The tests were drawn from environmental, information processing, neuropsychological, and microcomputer task batteries (64, 65). More than 140 performance measures were evaluated and documented in 90 reports (50). Because of the volume of this literature, a review focused on the utility of tasks is needed to enhance applicability. This report provides a synoptic evaluation of the human performance measures studied as part of the PETER Program.

Repeated Measures Applications

There are many situations in which it is useful to measure repeatedly human performance capabilities . These include following the time-course of performance in studies of vigilance, maturation, or environmental stress (75), and monitoring recovery from an injury (80). In addition, repeated measures are useful in evaluating the effectiveness of training (45) and in comparing the effects of various equipment configurations on human performance (88). The application of repeated measures spans the breadth of human performance experimentation.

Repeated measures experimentation is frequently favored in applied situations because it can be more efficient and economical than alternate approaches (110). When intertrial correlations are constant (i.e., differentially stable), the power of repeated measures analysis-of-variance increases with the magnitude of the correlations and considerable economy is realized (101). When two sets of measures have constant correlations, the power of differential (correlational) analyses may also be substantially increased by the use of correlated averages (35, 90) or more potently, by averaged correlations (11, 14, 32). However, economical use of subjects may provide the paramount rationale for repeated measures. This is true when there is a scarcity of qualified subjects or, more importantly, when there are hazards associated with the experimentation (15,25). Repeated measures designs permit the use of fewer subjects, but in addition, they minimize the total exposure time. Clearly, it is important to consider the task characteristics required for repeated measures applications before conducting research.

Criteria for Repeated Measures

Repeated measurements must possess certain characteristics in order to be meaningful, and to be easily and clearly interpretable (3,56,77). First, the measurements must represent a constant mixture of human performance capabilities on each trial of repeated measurement. In its simplest form, this requirement implies that the relative differences between subjects on the capability being measured remain constant across all trials of repeated measurement. This requirement for meaningful repeated measurements can be met objectively by showing that, apart from measurement errors, intertrial correlations are unchanging (differentially stable) and variances are homogeneous across baseline repetitions (9,57,77). Differential stability, in this context, provides assurance that the entity which is being measured is remaining constant (2). Stated technically, differential stability and constant variances make up the compound symmetry requirement of the variance-covariance matrix for simple repeated measures analysis of variance (110). Together, differential and variance stability are required for simplified analysis and interpretation.

The second requirement for meaningful and interpretable repeated measurements is that practice effects must be nil or predictable. In this regard, Lord and Novick (77) point out that repeated measurements may be useful if mean scores change by an additive constant from one trial to another. Campbell and Stanley (17), in their classic discussion, illustrate the principle that the additive constant should be the same from one trial to the next; the cumulative effect should have no more than a linear trend (preferably with near zero slope). Campbell and Stanley have also noted that nonlinear changes across repeated measurements impede or make impossible interpretation of effects of experimental interventions.

In sum, the statistical requirements for easily interpretable results of repeated measures include level or linearly increasing means, level variances, and differential stability.

PETER Paradigm

The PETER Program has focused largely upon determining when, if ever, practiced capability measures meet the criteria for repeated measures applications. In the typical evaluation procedure, a moderate number (10-25) of subjects were assessed daily for 15 days under baseline conditions at the same time of day. Also, massed practice effects were investigated in more abbreviated (3- to 10-day) studies in which multiple trials were given within a day (71,74).

A sequential strategy was employed in all studies to assess when means, variances, and intertrial correlations became stable (12,18). For the most part, this strategy involved dropping leading trials (these were usually daily scores) until an appropriate test statistic was conservatively nonsignificant ($p > 0.1$). For massed trials, within a day, the procedure was altered, on a case-by-case basis, to focus on trials not affected by massing effects. (e.g., first trials across days). In sum, the PETER paradigm was aimed at determining when, if ever, practiced tasks obtained mean, variance, and differential stability.

Subjects were U. S. Navy enlisted men, ages 18-28, who had volunteered for assignment to this laboratory as full-time research participants under provisions of informed consent (SECNAVINST 3900.39 Series and NAVMEDCOMINST 3900.5 Series). Subjects were selected for physical and other characteristics to participate in biodynamic research. They were intellectually typical of enlisted personnel (102).

Purpose

The purpose of this report is to provide an applications-oriented review of the performance measures evaluated as part of the PETER Program. Results for 112 measures are classified for their potential utility for the practioner. Discussion covers the application of the results and implications for past and future research.

METHOD

A survey was conducted and salient features were extracted from tasks studied in the PETER Program. Measures were categorized into four classes, depending upon their utility for repeated measures applications: Recommended, Acceptable-But-Redundant, Marginal, and Unacceptable.

Survey of Performance Measures

More than 140 performance measures were identified initially from documents listed in a recent bibliography of the PETER Program (50). Many tasks were excluded from consideration as they had been eliminated in the early stages of analysis, or were still at an early developmental level. The poor reliabilities and stabilities of difference, proportion, slope, and other derived measures eliminated many of them from consideration for repeated measures applications and discouraged complete documentation (10,22). Some computer mechanized tasks were not considered because they still required substantial development. These tasks frequently had less reliability than their paper-and-pencil counterparts, or had questionable construct validity (71,74,96). Other computer tasks which appear to have desirable metric qualities have been developed but are not in a sufficiently advanced stage to be included in this review (16). Overall, a total of 112 of the original 140 performance measures were finally judged adequate for complete reporting of the critical elements outlined in Table 1.

Mean, variance, and differential stability results for the 112 selected measures were evaluated for comparability before the features were extracted. This was necessary because statistical and interpretive methodology had evolved over the seven years of the PETER Program (12,13). Evaluations of differential stability, for example, were conducted by a half dozen approaches ranging from analysis based on graphical approaches (9) to analysis based on the work of Steiger (97,98). Where analyses were not comparable, data were reanalyzed by appropriate techniques (12). This was required, for example, where factor analysis was the method for establishing differential stability (58). Hence, the stability results were made comparable for the 112 measures *before salient features* were extracted.

Categorization of Measures

In the second stage of the investigation the 112 measures were categorized into the four groups: Recommended, Acceptable-But-Redundant, Marginal, and Unacceptable. This categorization was based upon joint consideration of task stability and task definition. This classification was designed as a guide for the selection of tasks for environmental and other repeated measures studies.

Recommended. Measures in this category were those that clearly obtained total stabilization and had an acceptable level of reliability efficiency (i.e., $r_{xx} \geq .50$, when normalized to a three-minute administration). This level of reliability was required for categorization as Recommended based upon earlier considerations of the statistical power of repeated measures designs (12).

Acceptable-But-Redundant. These measures had met the same requirements as those in the Recommended category, but had been found redundant by factor analysis or related studies of stabilized tasks. In addition to being redundant, these measures generally had slightly less reliability than their counterparts classified as Recommended.

Marginal. Marginal measures were distinguished by either instability of means or variances throughout practice, questionable differential stability, or less than a modicum of reliability efficiency ($.15 \leq r_{xx} < .50$). These measures usually had desirable features which were outweighed by faults.

Unacceptable. Measures in this category were characterized by either differential instability or weak reliability efficiency ($r_{xx} < .15$). This category contained an inordinate number of slope, difference, proportion, and other derived measures.

RESULTS

The tasks are categorized as Recommended, Acceptable-But-Redundant, Marginal, or Unacceptable in Tables 2 through 5. Definitions of the task features listed in the table headings are given in Table 1.

Recommended and Acceptable-But-Redundant

The Recommended and Acceptable-But-Redundant measures are summarized in Tables 2 and 3, respectively. Table 2 is made up of 30 measures of cognitive (17), perceptual (11), and motor (7) performance. (Note that Contrast Sensitivity constitutes five measures.) Table 3, which contains the Acceptable-But-Redundant, is made up of 15 measures which are primarily cognitive and perceptual. The scarcity of motor measures reflects an emphasis on factor-analytic and related differential studies of cognitive and perceptual tasks during the PETER Program. The Recommended and Acceptable-But-Redundant categories contain a wide range of tests of individual capabilities which we consider suitable for repeated measures research.

Marginal

Table 4 summarizes 35 measures which had one or more undesirable features and, therefore, could not be designated as totally suitable for repeated measures applications. Cognitive components are present in 20 measures. Major perceptual components are present in 27 measures, including 10 for Contrast Sensitivity. Fourteen measures, including 7 microprocessor-based games, have major motor components. Over the 35 measures, 4 of them are slope or difference scores. Flaws have been found in a broad range of performance measures.

Some of these tests could be of limited use in their present form. For example, otherwise flawed measures which became differentially stable with high reliability efficiences might be employed in purely differential correlational studies in which changes in the means and variances were of less interest. Other measures which may obtain total stability but had weak reliability efficiencies ($r_{xx} < .50$) might be considered for application if there were no other measure of that capability available. Extensive repetitions (more trials) would be required to insure power in cases where reliabilities are weak. However, before use of these measures, consideration should be given to task or scoring changes which could eliminate the undesirable features. Overall, while these Marginal tasks have some potential for application, first consideration should be given to making them suitable.

Unacceptable

Table 5 lists 32 measures found unsuitable for repeated measures applications in their present form. Thirteen of these measures have primarily cognitive components. Of the 17 measures having major perceptual components, 10 measures are summarized under the two entries for visual contrast sensitivity. The remaining four measures have major motor components. Ratio, slope, intercept, difference, and various derived scores make up 11 of the 32 measures categorized as Unacceptable.

DISCUSSION

The stability of 112 performance measures administered repeatedly under baseline conditions was reviewed. It was found that, although largely drawn from performance batteries, only 45 measures could be judged as Recommended or Acceptable-but-Redundant. Thus only about 40% of the well-practiced measures demonstrated total (mean, variance, and differential) stability. These and related findings provide a basis for the selection of tasks and pretest stabilization periods and will be discussed in this section. Methods of scoring, implications for the current environmental effects literature, and other findings will also be discussed.

Test Selection and Use

The results of the present review provide guidance for performance test selection and utilization. Table 2 delineates a range of 30 perceptual, cognitive, and motor measures which should be considered for repeated measures applications. Tables 3 to 5 outline 92 measures which cannot be recommended. In particular, Table 3 lists measures found suitable for repeated measures applications, but redundant with those in Table 2. Table 4 lists measures of questionable utility in their present format which should be considered for application only when no comparable measure can be found in the Recommended or Acceptable-But-Redundant categories. Substantial task development, to eliminate flaws, is recommended for measures in this category before their use. Table 5 lists measures found unsuitable for repeated measures use in their present format. In sum, measures suitable and unsuitable for repeated measures applications are identified in Tables 2 through 5. The researcher may consult these tables to determine the utility of a particular measure or the likely stability of a related one.

Table 2 provides selection and utilization information in addition to being an aggregation of fully suitable measures. Factor and domain information, in particular, may be used to identify subsets of me[illegible] for a particular application. For example, Guignard, Bittner, and Car[illegible] (47) used such an approach to identify five perceptual, cognitive, and mot[illegible] measures for use in an investigation of whole-body vibration. Reliability e[illegible]iciency data may be employed to select sensitive tasks from measure subse[illegible]. High reliability efficiencies provide for statistical power (20,101). For example, the approach of Guignard et al. (47) has been used to select a mini-battery for environmental applications. Table 6 characterizes this battery which contains tasks designed to assess left and right hemisphere functions, as well as fine perceptual motor and arm movement speed. The mini-battery assesses five measures with reliabilities above .85 in less than 10 minutes.

Prior to task selection, total stabilization time may be used in planning the amount of experimental practice time. Guignard et al. (47) used stabilization time information in planning their study. Anticipating the effects of massed practice on stabilization, Krause and Woldstad (74) allowed more practice than the minimum required for distributed practice. Altogether, the factor, domain, reliability, and stabilization information are an aid for selection and utilization of experimental tasks.

Scoring Methods

Analysis of the 112 measures indicated that derived scores frequently have undesirable properties (10,22,52). Specifically, none of the 15 difference, slope, or proportion scores may be seen in either the Recommended or Acceptable-But-Redundant category; while 45 of the 97 nonderived scores are classified in these categories ($\chi^2_{(1)} = 9.47$; $p < .005$). This association underestimates that across all 140 measures, derived scores made up a disproportionate number that were dropped early from consideration because of poor statistical characteristics. Overall, derived scores are associated with ratings of Marginal or Unacceptable.

During the present study, a combination of analytic and empirical evidence was uncovered which questions the use of difference-related scores.

Specifically, this report supports the analytic results of Cronbach and Furby (28) who found that difference scores tend to be unreliable and of questionable utility. For example, Harbeson, Krause, Kennedy, and Bittner (52) found that the Stroop interference score possessed low reliability. Moreover, this score was found to reflect a difference between two variables of virtually identical factor composition. Paralleling Chronbach and Furby, Carter and Krause (22) have demonstrated analytically that slope scores have properties similar to, if not isomorphic with, difference scores. In addition, they reported empirical slope score results which tended to exhibit low reliability and differential instability over a series of information processing tasks such as Short-Term Memory Scanning (99,100), and Letter Search (84). Similarily, Bittner (10) has demonstrated analytically the potential for undersirable properties with proportion-of-baselines and other ratios of random variables which are also difference-related. These properties were seen in earlier research (83,95) and indicate that often results using proportion of baseline may be artifactual (10). The present review suggests that the use of difference, slope, and proportional scores should be questioned.

The frequently undesirable properties of difference-related scores suggest a cautious empirical examination before they are used. Examination of theoretical models for individual subject derived scores may be recommended as a first step (10,22). As a second step, the methods for stability analysis described earlier (12) are also recommended after selection of an appropriate model. Evaluation the of the stability of difference-related scores is recommended to ensure meaningful experimentation.

Implications for the Environmental Literature

The finding that only 40% of the well-practiced tasks demonstrated total stability across repeated measurements brings into question the validity of that part of the performance literature based on repeated measures. Failure to meet the assumptions of total stability may be catastrophic. Nonlinear changes in means may render interpretation of intervention effects difficult, if not impossible (17). In addition, a failure to obtain joint variance and differential stability implies seriously distorted statistical tests for effects and, consequently, misleading evidence as to the presence of such effects (91,110). Examining mean and variance stability, graphically and otherwise, is a good first step before initiating investigations. Unfortunately, stability of means and variances does not imply differential stability (80). Examining only means and variances may result in failure to identify changes in the nature of the construct which is being measured; differential changes make the meaningful interpretation of the results virtually impossible. Failure to attend to task stability may be a source of the difficulties in meta-analyses of environmental literature (40). It is concluded that the validity of much of the environmental research literature could be questioned on the grounds of possible instability of repeated measures.

The results of this review also support the validity of part of the environmental literature. Many investigators have used one of the measures identified in the Recommended or Acceptable-But-Redundant categories in the present review and have practiced subjects sufficiently to have obtained stability. Baddeley's (4) Grammatical Reasoning test, for example, has been

TABLE 3: ACCEPTABLE-BUT-REDUNDANT

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE	TOT STAB TIME IN MINUTES (DIFF)	RELIAB	EFFIC	3 MIN	REFERENCES
ARITHMETIC COMPUTATION	NUMBER FACILITY (N) (EKSTROM ET AL., 1976)	C	10	G	90(<10)	0.83			SEALES ET AL. (1980)
ARITHMETIC: NUMBER FACILITY	NUMBER FACILITY (N) (EKSTROM ET AL., 1976)	C	3	G	27(27)	0.83			BITTNER ET AL.(1983); MORAN & MEFFERD (1959)
CHOICE REACTION TIME: 2-CHOICE	CHOICE REACTION TIME (DONDERS, 1868)	P	5.0	I	35(35)	0.51			KRAUSE & BITTNER (1982); TEICHNER & KREBS (1974)
GRAPHEMIC AND PHONEMIC ANALYSIS: SENSE/HOMOPHONE	VISUAL OR GRAPHEMIC ENCODING (BARON & MCKILLOP, 1975)	C	8	G	40(40)	0.66			HARBESON, KENNEDY, ET AL. (1982); BARON & MCKILLOP (1975); ROSE & FERNANDES (1977)
GRAPHEMIC AND PHONEMIC ANALYSIS:HOMOPHONE /NONSENSE	ACOUSTIC OR PHONEMIC ENCODING (BARON & MCKILLOP, 1975)	C	8	G	72(72)	0.73			HARBESON, KENNEDY, ET AL. (1982); BARON & MCKILLOP (1975); ROSE & FERNANDES (1977)
LETTER CLASSIFICATION: PHYSICAL MATCH	PATTERN MATCHING (POSNER & MITCHELL, 1967)	P	12	G	108(108)	0.52			HARBESON, KENNEDY, ET AL. (1982); POSNER & MITCHELL (1967); ROSE & FERNANDES (1977)
LETTER SEARCH: TIME PER CORR. ITEM	VISUAL SEARCH (NEISSER ET AL., 1963)	P	3	G	27(27)	0.87			CARTER & KRAUSE (1983); CARTER & SBISA (1982) SHANNON ET AL. (IN PRESS)
MINNESOTA RATE MANIPULATION: PLACING	MANUAL DEXTERITY (FLEISHMAN & ELLISON, 1962)	M	3-5	I	42(42	0.61			CARTER, STONE, & BITTNER (1982); SCHOENFELDT (1972)
NUMBER COMPARISON	PERCEPTUAL SPEED (P) (EKSTROM ET AL., 1976)	P	3	G	27(9)	0.84			BITTNER ET AL. (1983); CARTER & SBISA (1982)
PATTERN RECOGNITION TIME PER CORRECT ITEM	PATTERN RECOGNITION (FITTS, WEINSTEIN, RAPPAPORT, ET AL., 1956)	P	2	G	20(20)	0.76			CARTER & SBISA (1982); CARTER & KRAUSE (1983)
PURDUE PEGBOARD	FINE FINGER DEXTERITY (TIFFIN, 1968)	P M	2	I	42(42)	0.90			KRAUSE & WOLDSTAD (1983); TIFFIN (1968)
RANDOM FIELD NUMBER SEARCH: TIME PER CORRECT ITEM	VISUAL SEARCH	P	5	G	35(35)	0.55			SHANNON ET AL. (IN PRESS) CARTER & SBISA (1982)
SPEED OF CLOSURE	CLOSURE, VERBAL (CV) (EKSTROM ET AL., 1976)	P	2.5	G	28(25)	0.80			BITTNER ET AL. (1983); MORAN & MEFFERD (1959)
STROOP: BLACK & WHITE WORDS (BW)	PERCEPTUAL SPEED (JENSEN & ROHWER, 1966)	P	0.5	G	1.5(.5)	0.96			HARBESON, KRAUSE, ET AL. (1982)
STROOP: COLOR BLOCKS (CB)	MIXED	P	0.5	G	3.5(3.5)	0.98			HARBESON, KRAUSE, ET AL. (1982)

TABLE 2: RECOMMENDED (CONTINUED)

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE	ADMIN	TOT STAB TIME IN MINUTES (DIFF)	RELIAB EFFIC 3 MIN	REFERENCES
MANIKIN TEST: LOG. LATENCY	SPATIAL TRANSFORMATION (EGAN, 1978)	P	7	I		14(14)	0.79	CARTER & WOLDSTAD (IN PRESS); READER, BENEL, & RAHE (1981)
MINNESOTA RATE OF MANIPULATION: TURNING	MANUAL DEXTERITY (FLEISHMAN & ELLISON, 1962)	M	2-4	I		10(10)	0.64	CARTER, STONE, & BITTNER (1982); SCHOENFELDT (1972)
PATTERN COMPARISON: NUMBER CORRECT MINUS NUMBER INCORRECT	SPATIAL ABILITY (KLEIN & ARMITAGE, 1979)	P	2	G		18(18)	0.93	SHANNON, CARTER, & BOUDREAU (IN PRESS); KLEIN & ARMITAGE (1979); CARTER & SBISA (1982)
PERCEPTUAL SPEED	PERCEPTUAL SPEED (PS) (EKSTROM ET AL., 1976)	P	2.5	G		23(15)	0.86	BITTNER ET AL. (1983); MORAN & MEFFERD (1959)
SEARCH FOR TYPOS IN PROSE: MEDIAN DETECTION TIME	READING SPEED	P	6	I		54(54)	0.65	SHANNON ET AL. (IN PRESS); CARTER & KRAUSE (1983)
SPOKE CONTROL (C) TASK	SPEED ARM MOVEMENT (FLEISHMAN & ELLISION, 1962)	M	0.67 APPROX	G		1(1)	0.95	BITTNER, LUNDY, KENNEDY, & HARBESON (1982)
STERNBERG ITEM RECOGNITION: POSITIVE SET 1	SHORT TERM MEMORY SCAN (STERNBERG, 1966, 1975)	C	3	I		18(18)	0.70	CARTER, KENNEDY, BITTNER, & KRAUSE (1980); STERNBERG (1969, 1975)
STERNBERG ITEM RECOGNITION: POSITIVE SET 4	SHORT-TERM MEMORY SCAN (STERNBERG, 1966, 1975)	C	3	I		15(9)	0.80	CARTER ET AL. (1980); CARTER & KRAUSE (1983); STERNBERG (1969, 1975)
STROOP: COLOR WORDS (CW)	MIXED	C P	0.5	G		1.5(1.5)	0.97	HARBESON, KRAUSE, KENNEDY, & BITTNER (1982)
TRACKING: CRITICAL	TRACKING, CRITICAL (JEX ET AL., 1966)	P M	1	I		100(100)	0.60	DAMOS ET AL. (1984); JEX ET AL. (1966)
TRACKING: DUAL CRITICAL	TRACKING, CRITICAL & DUAL FACTOR? (DAMOS, BITTNER, KENNEDY, & HARBESON, 1981)	P M	1	I		100(100)	0.50	DAMOS ET AL. (1981)
VISUAL CONTRAST SENSITIVITY: METHOD OF INCREASING CONTRAST	CONTRAST SENSITIVITY FUNCTION: 1, 2, 4, 8, 16 cpd (GINSBURG & EVANS, 1982)	P P P P P	3 3 3 3 3	I I I I I		<1(<1) <1(<1) <1(<1) <1(<1) <1(<1)	0.51 0.52 0.74 0.75 0.53	GINSBURG, BITTNER, KENNEDY, HARBESON (1983); GINSBURG & EVANS (1982)
WORD FLUENCY	WORD FLUENCY (FW) (EKSTROM ET AL., 1976)	C	3	G		<1(<1)	0.79	CARTER, CURLEY, & STYER (IN REVIEW)

TABLE 2: RECOMMENDED*

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE ADMIN	TOT STAB TIME IN MINUTES (DIFF)	RELIAB	EFFIC 3 MIN	REFERENCES
AIMING	AIMING: FINE EYE-HAND COORDINATION (FLEISHMAN & ELLISON, 1962)	P M	2	G	30(30)	0.87		KRAUSE & WOLDSTAD (1983); FLEISHMAN & ELLISON (1962)
ARITHMETIC: VERTICAL ADDITION	NUMBER FACILITY (N) (EKSTROM, FRENCH, HARMON, & DERMEN, 1976)	C	4	G	48(8)	0.90		BITTNER, CARTER, KRAUSE, KENNEDY, & HARBESON (1983); CARTER & SBISA (1982)
ASSOCIATIVE MEMORY: NUMBER CORR: LIST 1	ASSOCIATIVE MEMORY (MA) (EKSTROM ET AL., 1976)	C	2.5	G	20(20)	0.65		KRAUSE & KENNEDY, 1980 CARTER & KRAUSE (1983); UNDERWOOD BORUCH & MALMI (1977)
ATARI® AIR COMBAT MANEUVERING	PURSUIT TRACKING (KENNEDY, BITTNER & JONES, 1981)	P M	2.25	I	135(135)	0.63		JONES, KENNEDY, & BITTNER (1981); KENNEDY, BITTNER, HARBESON, & JONES (1982)
ATARI® ANTIAIRCRAFT	UNKNOWN	P M	2.25	I	126(126)	0.67		JONES & KENNEDY (1983) WITH ADAPTATIONS
CHOICE REACTION TIME: 1-CHOICE	SIMPLE REACTION TIME (DONDERS, 1868)	P	5.0	I	35(35)	0.58		KRAUSE & BITTNER (1982); TEICHNER & KREBS (1974)
CHOICE REACTION TIME: 4-CHOICE	CHOICE REACTION TIME (DONDERS, 1868)	P	5.0	I	50(50)	0.80		KRAUSE & BITTNER (1982); TEICHNER & KREBS (1974)
CODE SUBSTITUTION	MEMORY ASSOC.(MA) PERCEPTUAL SPEED (P)(EKSTROM ET AL., 1976)	C P	2.0	G	16(16)	0.84		PEPPER, KENNEDY, BITTNER, & WIKER (1980); WECHSLER (1981)
FLEXIBILITY OF CLOSURE	CLOSURE, FLEXI-BILITY OF (CF) (EKSTROM ET AL., 1976)	P	3	G	9(9)	0.88		BITTNER ET AL. (1983); MORAN & MEFFERD (1959)
GRAMMATICAL REASONING	REASONING, LOGI-CAL (RL) (EKSTROM ET AL., 1976)	C	1.5	G	18(18)	0.93		BITTNER ET AL. (1983); CARTER, KENNEDY, & BITTNER (1981); BADDELEY (1968)
GRAPHEMIC AND PHONEMIC ANAL-YSIS: SENSE/ NONSENSE	READING SPEED (BARON & MCKILLOP, 1975)	C	8	G	16(16)	0.66		HARBESON, KENNEDY, KRAUSE, & BITTNER (1982); BARON & MCKILLOP (1973); ROSE & FERNANDES (1977)
LETTER CLASS-IFICATION: NAME	RETRIEVAL FROM LTM & MATCHING (POSNER & MITCHELL, 1967)	C	12	G	84(84)	0.55		HARBESON, KENNEDY, ET AL. (1982); POSNER & MITCHELL (1967); ROSE & FERNANDES (1977)
LETTER CLASS-IFICATION: CATEGORY	RETRIEVAL FROM LTM & MATCHING (POSNER & MITCHELL, 1967	C	11	G	121(121)	0.69		HARBESON, KENNEDY, ET AL. (1982); POSNER & MITCHELL (1967); ROSE & FERNANDES (1977)

*Continued on next page.

TABLE 1. DEFINITIONS OF TASK FEATURES

FEATURE (Abbreviations used in tables)	DEFINITION
NAME	Name of the task or measure as used in the literature.
FACTOR	The factor(s) assessed by the measure as identified in the literature or by judgments of the authors.
DOMAIN	Characterization of the domain(s) of assessment of the capability as cognitive, perceptual (including sensory), or motor.
ADMINISTRATION TIME IN MINUTES (ADMIN TIME)	The typical testing time for a measure; this includes all testing time required to obtain a score (e.g., components of a derived score)
TYPE OF ADMINSTRATION (TYPE ADMIN)	Identification of task as individually (I) or group (G) administered.
TOTAL STABILIZATION TIME IN MINUTES (DIFFERENTIAL)	The total stabilization time is the amount of elapsed experimental time (whether massed or distributed) required for mean, variance, and differential (correlational) stabilization. (The amount of elapsed practice time required for Differential Stabilization alone is in parentheses).
RELIABILITY EFFICIENCY (3 minutes)	The differentially stabilized reliability normalized to a 3 minute administration. Normalization to 3 minutes was by the Spearman-Brown Equation (Bittner & Carter, 1981; Winer, 1971).
REFERENCES	Cited in order are the relevant stability study, the original source of the measure, and occasionally other significant references.

106. Vroon PA. Sequential estimations of time. Acta Psychol. 1976; 40:457-487.

107. Webb WB, Levy CM. Age, sleep deprevation, and performance. Psychophysiology, 1982; 19:272-291.

108. Wechsler D. WAIS-R Manual: Wechsler Adult Intellegence Scale-Revised. New York, NY: The Psychological Corporation, 1981.

109. Wiker SF, Kennedy RS, Pepper RL. Performance Evaluation Tests for Environmental Research (PETER): Navigational plotting. Aviat. Space Environ. Med. 1983; 54:144-149.

110. Winer BJ. Statistical principles in experimental design. 2nd ed. New York, NY: McGraw-Hill, 1971.

111. Wonderlic EF. Wonderlic Personnel Test Manual. Northfield, IL: Wonderlic, 1978.

112. Zelkind I, Sprung J. Time research: 1172 studies. Metuchen, NJ: Scarecrow Press, 1974.

92. Schoenfeldt LF. Minnesota Rate of Manipulation Test, 1969 edition. In Buros OK, ed. The seventh mental measurements yearbook. Highland Park, NJ: Gryphon Press, 1972.

93. Shannon RH, Carter RC, Boudreau YA. A systematic approach to battery development and testing within unusual environments. In: Guignard JC, Harbeson MM, eds. Proceedings of the International Workshop on Research Methods in Human Motion and Vibration Studies. New Orleans, LA: Naval Biodynamics Laboratory, in press.

94. Shannon RH, Krause M, Irons RC. Attribute requirements for a simulated flight scenario microcomputer test. Research Report NBDL-82R004, NTIS No. AD A115676. New Orleans, LA: Naval Biodynamics Laboratory, Jan. 1982.

95. Seales DM, Kennedy RS, Bittner AC Jr. Development of a Performance Evaluation Test for Environmental Research (PETER): Arithmetic computation. Percept. Mot. Skills 1980; 51:1023-1031.

96. Smith MG, Krause M, Kennedy RS, Bittner, AC Jr, Harbeson, MM. Performance testing with microprocessors: Mechanization is not implementation. Proceedings of the 27th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1983: 674-678.

97. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol. Bull. 1980; 87:245-251.

98. Steiger JH. Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. Multivar. Behav. Res. 1980; 15:335-352.

99. Sternberg S. High speed scanning in human memory. Science 1966; 153:652-654.

100. Sternberg S. Memory scanning: New findings and current controversies. Q. J. Exp. Psychol. 1975; 27:1-32.

101. Sutcliffe JP. On the relationship of reliability to statistical power. Psychol. Bull. 1980; 88:509-515.

102. Thomas DJ, Majewski PL, Ewing CL, Gilbert NS. Medical qualification procedures for hazardous-duty aeromedical research. AGARD Conference Proceedings No. 231. Neuilly-Sur-Seine, France: NATO/AGARD; 1978; Section A-3:1-13.

103. Tiffin J. Manual for Purdue Pegboard. Chicago, IL: Science Research Associates, 1968.

104. Teichner WH, Krebs MJ. Laws of visual choice reaction time. Psychol. Rev. 1974; 81:75-98.

105. Underwood BJ, Boruch RF, Malmi RA. The composition of episodic memory. ONR Contract No. N00014-76-C-0270, NTIS No. AD A040696. Evanston, IL: Northwestern University, May 1977.

79. McCafferty DB, Bittner AC Jr, Carter RC. Performance Evaluation Tests for Environmental Research (PETER): Auditory digit span task. Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1980: 330-334. Also, Research Report No. NBDL-80R008, NTIS No. AD A111296. New Orleans, LA: Naval Biodynamics Laboratory, July 1981: 29-33.

80. McCauley ME, Kennedy RS, Bittner AC Jr. Development of a Performance Evaluation Test for Environmental Research (PETER): Time estimation test. Percept. Mot. Skills 1980; 51:655-665.

81. Meyer DE, Schvaneveldt RW, Ruddy MG. Functions of graphemic and phonemic codes in visual recognition. Memory and Cognition, 1974 2:309-321.

82. Moran LJ, Mefferd RB. Repetitive psychometric measures. Psychol. Rep. 1959; 5:269-275.

83. Morrissey SJ, Bittner AC Jr. Effects of vibration on humans: Performance decrements and limits. TP-75-47. Point Mugu, CA: Pacific Missile Test Center, 1975.

84. Neisser U, Novick R, Lazar R. Searching for ten targets simultaneously. Percept. Mot. Skills 1963; 17:955-961.

85. Pepper RL, Kennedy RS, Bittner AC Jr, Wiker SF. Performance Evaluation Tests for Environmental Research (PETER): Code substitution test. Proceedings of the 7th Psychology in the DoD Symposium. USAFA-TR-80-12. Colorado Springs, CO: USAF Academy, Apr. 1980: 451-457. Also, Research Report No. NBDL-80R008, NTIS No. AD A111296. New Orleans, LA: Naval Biodynamics Laboratory, July 1981: 13-19.

86. Posner MI, Mitchell RF. Chronometric analysis of classification. Psychol. Rev. 1967; 74:392-409.

87. Reader DC, Benel RA, Rahe AJ. Evaluation of a manikin psychomotor task. Report No. SAM-TR-81-10, NTIS No. AD A100966. Brooks Air Force Base, TX: USAF School of Aerospace Medicine, 1981.

88. Roscoe SN. Aviation psychology. Ames, IA: Iowa State University Press, 1980.

89. Rose AM, Fernandes K. An information processing approach to performance Assessment: I Experimental investigation of an information processing performance battery. Report No. AIR-58500-TR, NTIS No. AD A047299. Washington, DC: American Institutes for Research, Nov. 1977.

90. Rushton JP, Brainerd CJ, Pressley M. Behavioral development and construct validity: The principle of aggregation. Psychol. Bull. 1983; 94:18-38.

91. Scheffe, H. The analysis of variance. New York, NY: Wiley; 1959.

67. Kennedy RS, Bittner AC Jr, Jones MB. Video game and conventional tracking. Percept. Mot. Skills 1981; 53:310.

68. Kennedy RS, Bruns RA. Some practical considerations for performance testing in exotic environments. AGARD Conference Proceedings No. 181. Neuilly-sur-Seine, France: NATO/AGARD, 1975; Section C-4:1-6.

69. Kennedy RS, Harbeson MM. A retrospective view of the Performance Evaluation Tests for Environmental Research (PETER) Program. In: Guignard JC, Harbeson MM, eds. Proceedings of the International Workshop on Research Methods in Human Motion and Vibration Studies. New Orleans, LA: Naval Biodynamics Laboratory, in press.

70. Klein R, Armitage R. Rhythms in human performance: 1 1/2-hour oscillations in cognitive style. Science, 1979; 204:1326-1328.

71. Krause M. Paper-and-pencil and computerized performance tests: Does the medium make a difference? New Orleans, LA: Naval Biodynamics Laboratory, in preparation.

72. Krause M, Bittner AC Jr. Repeated measures on a choice reaction time task. Research Report No. NBDL-82R006, NTIS No. AD A121904. New Orleans, LA: Naval Biodynamics Laboratory, Nov. 1982.

73. Krause M, Kennedy RS. Development of Performance Evaluation Tests for Environmental Research (PETER): Interference susceptibility test. Proceedings of the 7th Psychology in the DoD Symposium. USAFA-TR-80-12. Colorado Springs, CO: USAF Academy, Apr. 1980: 459-464. Also, Research Report No. NBDL-80R008, NTIS No. AD A111296. New Orleans, LA: Naval Biodynamics Laboratory, July 1981: 41-46.

74. Krause M, Woldstad JC. Massed practice: does it change the statistical properties of performance tests? Research Report No. NBDL-83R005, NTIS No. AD A139338. New Orleans, LA: Naval Biodynamics Laboratory, Apr. 1983.

75. Lewis VM, Baddeley AD. Cognitive performance, sleep quality, and mood during deep oxyhelium diving. Ergonomics, 1981; 24:773-793.

76. Logie RH, Baddeley AD. A trimix saturation dive to 660 m: Studies of cognitive performance, mood and sleep quality. Ergonomics, 1983; 26:359-374.

77. Lord FM, Novick MR. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

78. Mackaman SM, Bittner AC Jr, Harbeson MM, Kennedy RS, Stone D A. Performance Evaluation Tests for Environmental Research (PETER): Wonderlic Personnel Inventory. Psychol. Rep. 1982; 51:635-644.

79. McCafferty DB, Bittner AC Jr, Carter RC. Performance Evaluation Tests for Environmental Research (PETER): Auditory digit span task. Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1980: 330-334. Also,

56. Jones, MB. Individual differences. In: Singer RN, ed. The psychomotor domain. Philadelphia, PA: Lea & Febiger, 1972: 107-132.

57. Jones MB. Stabilization and task definition in a performance test battery. Final Report on Contract No. N0023-79-M-5089, Monograph No. NBDL-M001, NTIS No. AD A099987. New Orleans, LA: Naval Biodynamics Laboratory, Oct. 1980.

58. Jones MB, Kennedy RS. Stabilization with practice: Theory and , assessment. Unpublished manuscript, available from second author, Essex Corp., 1040 Woodcock Rd., Suite 227, Orlando, FL 32803, 1983.

59. Jones MB, Kennedy RS, Bittner AC Jr. Video games and convergence of divergence with practice. Proceedings of the 7th Psychology in the DoD Symposium. USAFA-TR-80-12. Colorado Springs, CO: USAF Academy, Apr. 1980: 465-469. Also, Research Report No. NBDL-81R010, NTIS No. AD A111086. New Orleans, LA: Naval Biodynamics Laboratory, Sept. 1981: 1-5.

60. Jones MB, Kennedy RS, Bittner AC Jr. A video game for performance testing. Am. J. Psychol. 1981; 94:143-152.

61. Kennedy RS, Andrews DA, Carter RC. Performance Evaluation Tests for Environmental Research (PETER): A microcomputer game as a memory test. Preprints of the 52nd Annual Scientific Meeting of the Aerospace Medical Association. Washington, DC: Aerospace Medical Association, 1981: 240-241.

62. Kennedy RS, Bittner AC Jr. The development of a Navy Performance Evaluation Test for Environmental Research (PETER). In: Pope LT, Meister D, eds. Productivity Enhancement: Personnel Performance Assessment in Navy Systems. San Diego, CA: Navy Personnel Research & Development Center, Oct. 1977. Also, Research Report No. NBDL-80R004, NTIS No. AD A111180. New Orleans, LA: Naval Biodynamics Laboratory, Nov. 1981: 22-37.

63. Kennedy RS, Bittner AC Jr. Performance Evaluation Tests for Environmental Research (PETER): Complex counting. Aviat. Space Environ. Med. 1980; 51:142-144.

64. Kennedy RS, Bittner AC Jr, Carter RC, Krause M, Harbeson MM, McCafferty DB, Pepper RL, Wiker SF. Performance Evaluation Tests for Environmental Research (PETER): Collected papers. Research Report No. NBDL-80R008, NTIS No. AD A111296. New Orleans, LA: Naval Biodynamics Laboratory, July 1981.

65. Kennedy RS, Bittner AC Jr, Harbeson MM, Jones MB. Perspectives in Performance Evaluation Tests for Environmental Research (PETER): Collected papers. Research Report No. NBDL-80R004, NTIS No. AD A111180. New Orleans, LA: Naval Biodynamics Laboratory, Nov. 1981.

66. Kennedy RS, Bittner AC Jr, Harbeson MM, Jones MB. Television-computer games: A "new look" in performance testing. Aviat. Space Environ. Med. 1982; 53:49-53.

44. Ginsburg AP, Evans DW. Contrast sensitivity predicts pilot's performance in aircraft simulators. Preprints of the 53rd Annual Meeting of the Aerospace Medical Association. Washington, DC: Aerospace Medical Association, 1982: 212-213.

45. Goldstein IL. Training: Program development and evaluation. Monterey, CA: Brooks/Cole, 1974.

46. Graybiel A, Kennedy RS, Knoblock EC, Guedry FE Jr, Mertz W, McLeod ME, Colehour JL. Miller EF II, Fregly AR. Effects of exposure to a rotating environment (10 rpm) on four aviators for a period of twelve days. Aerospace Med. 1965; 36:733-754.

47. Guignard JC, Bittner AC Jr, & Carter RC. Methodological investigation of vibration effects on performance of three tasks. Journal of Low Frequency Noise and Vibration, 1982, 1, 12-18.

48. Guignard JC, Bittner AC Jr, Einbender SW, Kennedy RS. Performance Evaluation Tests for Environmental Research (PETER): Landolt C reading test. Proceedings of the 24th Annual Meeting of the Human Factors Society, Santa Monica, CA.: Human Factors Society, 1980: 335-339.

49. Guignard JC, Landrum GJ, Reardon RE. Experimental Evaluation of Human Long-Term Vibration Exposures Permitted by the Current International Standard ISO 2631, 1974. Preprints of the 47th Annual Scientific Meeting of the Aerospace Medical Association. Washington, DC: Aerospace Medical Association, 1976.

50. Harbeson MM, Bittner AC Jr, Kennedy RS, Carter RC, Krause M. Performance Evaluation Tests for Environmental Research (PETER): Bibliography. Percept. Mot. Skills 1983; 57:283-293.

51. Harbeson MM, Kennedy RS, Krause M, Bittner AC Jr. Repeated measures of information processing. Proceedings of the 26th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1982: 818-822.

52. Harbeson MM, Krause M, Kennedy RS, Bittner AC Jr. The Stroop as a Performance Evaluation Test for Environmental Research. J. Psychol. 1982; 111:223-233.

53. Harbeson MM, Krause M, Kennedy RS. The comparison of memory tests for environmental research. Proceedings of the 24th Annual Meeting of the Human Factors Society, Santa Monica, CA: Human Factors Society, 1980: 349-353. Also, Research Report No. NBDL-80R008, NTIS No. AD A111296. New Orleans, LA: Naval Biodynamics Laboratory, July 1981: 34-40.

54. Jensen AR, Rohwer WD. The Stroop Color-Word Test: A review. Acta Psychol. 1966; 25:36-93.

55. Jex HR, McDonnell JD, Phatak AV. A "critical" tracking task for manual control research. IEEE Trans. Hum. Factors Electronics, 1966; HFE-7:138-145.

28. Cronbach LJ, Furby L. How should we measure change - or should we? Psychol. Bull. 1970; 74:68-70.

29. Damos DL, Bittner AC Jr, Kennedy RS, Harbeson MM. Effects of extended practice on dual-task tracking performance. Hum. Factors 1981; 23:627-631.

30. Damos DL, Bittner AC Jr, Kennedy RS, Harbeson MM. Krause M. Performance Evaluation Test for Environmental Research (PETER): Critical tracking test. Percept. Mot. Skills 1984; 58:564-573.

31. Donders FC. Die Schnelligkeit Psychischer Processe. Archiv fur Anatomie und Physiologie und Vissenschaftliche Medizin 1868; 657-681.

32. Dunlap WP, Jones MB, Bittner AC Jr. Average correlations versus correlated averages. Bull. Psychonom. Soc. 1983; 21:213-216.

33. Egan DE. Characterizing spatial ability: Different mental processes reflected in accuracy and latency scores. Research Report No. 1250. Pensacola, FL: Naval Aerospace Medical Research Laboratory, Aug. 1978.

34. Ekstrom RB, French JW, Harmon HH, Dermen D. Manual for kit of factor referenced cognitive tests. Princeton, NJ: Educational Testing Service, 1976.

35. Epstein S. The stability of behavior: I. On predicting most of the people much of the time. J. Pers. Soc. Psychol. 1979; 37:1097-1126.

36. Fernandes K, Rose AM. An information processing approach to performance assessment: An investigation of encoding and retrieval processes in memory. Tech. Report AIR-58500-11/78 TR. Washington, DC: American Institutes for Research, Nov. 1978.

37. Forrester WE. Publication trends in human learning and memory: 1962-1982. Bull. Psychonom. Soc. 1984; 22:92-94.

38. Fitts PM, Weinstein M, Rappaport M, Anderson N, Leonard J A. Stimulus correlates of visual pattern recognition: A probability approach. J. Exp. Psychol. 1956; 51:19-24.

39. Fleishman EA, Ellison GD. A factor analysis of fine manipulative tests. J. Appl. Psychol. 1962; 46:96-105.

40. Gawron VJ. Noise: Effect and aftereffect. Ergonomics 1984; 27:5-18.

41. Gebelli N. Phantoms Five®. Sacramento, CA: Sirius Software, 1980.

42. Gibson TR. Effects of hypocapnea on psychomotor and intellectual performance. Aviat. Space Environ. Med. 1978; 49:943-946.

43. Ginsburg AP, Bittner AC Jr, Kennedy RS, Harbeson MM. A methodological investigation of three psychophysical techniques for rapid measurement of contrast sensitivity. Proceedings of the 27th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1983: 264-268.

14. Bittner AC Jr, Dunlap WP, Jones MB. Averaged correlations with differentially-stable variables: Fewer subjects required for repeated measures. Proceedings of the 26th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1982: 349-353.

15. Bittner AC Jr, Lundy NC, Kennedy RS, Harbeson MM. Performance Evaluation Tests for Environmental Research (PETER): Spoke tasks. Percept. Mot. Skills 1982; 54:1319-1331.

16. Bittner AC Jr, Smith MG, Kennedy RS, Staley CF, Harbeson, MM. Automated Portable Test (APT) system: Overview and prospects. Presented at the 14th Annual Meeting of the Society for Computers in Psychology, San Antonio, TX, Nov. 1984. Manuscript available from authors.

17. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally, 1963.

18. Carter RC, Bittner AC Jr. Jackknife for variance analysis of multifactor experiments. Research Report No. NBDL-82R013, NTIS No. AD A121760. New Orleans, LA: Naval Biodynamics Laboratory, May 1982.

19. Carter RC, Curley M, Styer D. Repeated measurements of divers' word fluency. Manuscript submitted for publication.

20. Carter RC, Kennedy RS, Bittner AC Jr. Grammatical reasoning: A stable performance yardstick. Hum. Factors 1981; 23:587-591.

21. Carter RC, Kennedy RS, Bittner AC Jr, Krause M. Item recognition as a Performance Evaluation Test for Environmental Research. Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1980: 340-344. Also, Research Report No. NBDL-81R008, NTIS No. AD A111296. New Orleans, LA: Naval Biodynamics Laboratory, July 1981: 47-50.

22. Carter RC, Krause M. Reliability of slope scores for individuals. Research Report No. NBDL-83R003, NTIS No. AD A130252. New Orleans, LA: Naval Biodynamics Laboratory, Apr. 1983.

23. Carter RC, Sbisa HE. Human performance tests for repeated measurements: alternate forms of eight tests by computer. Research Report No. NBDL-82R003, NTIS No. AD A115021. New Orleans, LA: Naval Biodynamics Laboratory, Jan. 1982.

25. Carter RC, Stone DA, Bittner AC Jr. Repeated measurements of manual dexterity: Applications and support of the two-process theory. Ergonomics 1982; 25:829-838.

26. Carter RC, Woldstad JC. Repeated Measurements of spatial ability with the Manikin Test. Hum. Factors, in press.

27. Collins AM, Quillian, MR. Retrieval time from semantic memory. J. Verbal Learn. Verbal Behav. 1969; 8:240-247.

REFERENCES

1. Alluisi EA. Sustained performance. In: Bilodeau EA, Bilodeau IMcD, ed. Principles of skill acquisition. New York, NY: Academic Press, 1969: 50-101.

2. Alvares KM, Hulin CL. Two explanations of temporal changes in ability-skill relationships: a literature review and theoretical analysis. Hum. Factors 1972; 14:295-308.

3. American Psychological Association. Standards for educational and psychological tests. Washington, DC: American Psychological Association, 1974.

4. Baddeley AD. A three minute reasoning test based on grammatical transformation. Psychonom. Sci. 1968; 10:341-342.

5. Baddeley AD, De Figuerido JW, Curtis, JWH, Williams AN Nitrogen narcosis and performance under water. Ergonomics 1968; 11:157-164.

6. Baron J. Phonemic stage not necessary for reading. Q. J. Exp. Psychol. 1973; 25:241-246.

7. Baron J, McKillop BJ. Individual differences in speed of phonemic analysis, visual analysis, and reading. Acta Psychol. 1975; 39:91-96.

8. Bell E, Jurek K, Wilson T. Hand skill measurement: a gauge for treatment. Am. J. Occup. Treat. 1976; 30:80-86.

9. Bittner AC Jr. Statistical tests for differential stability. Proceedings of the 23rd Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, 1979: 541-545. Also, Research Report No. NBDL-81R010, NTIS No. AD A111086. New Orleans, LA: Naval Biodynamics Laboratory, Sept. 1981: 10-14.

10. Bittner AC Jr. Use of proportion-of-baseline measures in stress research. In: Salvendy G, Smith MJ, eds. Machine pacing and occupational stress. London: Taylor & Francis, 1981: 177-183.

11. Bittner AC Jr. Averaged correlations between parallel measures: reliability estimation. Proceedings of the 8th Psychology in the DoD Symposium (USAFA-TR-82-10). Colorado Springs, CO: USAF Academy, Apr. 1982: 321-327.

12. Bittner AC Jr, Carter RC. Repeated measures of human performance: A bag of research tools. Research Report No. NBDL-81R011, NTIS No. AD A113954. New Orleans, LA: Naval Biodynamics Laboratory, 1981.

13. Bittner AC Jr, Carter RC, Krause M, Kennedy RS, Harbeson MM. Performance Evaluation Tests for Environmental Research (PETER): Moran and Computer Batteries. Aviat. Space Environ. Med. 1983; 54:923-928.

employed in a substantial number of environmental investigations (5,42,75,76, 107). Critical tracking (55) has also been extensively employed (30). Practice levels routinely recommended by Jex (personal communication, 1983) exceed those recommended in this report by 50%. Investigations which have appropriately used stable measures provide a firm foundation for the understanding of environmental effects.

CONCLUSIONS AND RECOMMENDATIONS

Four conclusions and associated recommendations emerged from the present research:

1. The tables presented in this report are a guide for the selection and application of performance tests in environmental research.

2. Difference, slope, and ratio scores frequently possess undesirable psychometric properties, and their cautious empirical examination is recommended before application.

3. The literature on performance changes due to environmental factors should be reviewed in terms of stability or instability of measurements.

4. The evaluation of the psychometric stability of performance measures under baseline conditions provides a foundation for environmental research applications using repeated measurements.

TABLE 4. MARGINAL*

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE	ADMIN	TOT STAB TIME IN MINUTES (DIFF)	RELIAB	EFFIC	3 MIN	REFERENCES
ARITHMETIC: HORIZONTAL ADDITION	NUMBER FACILITY (N) (EKSTROM ET AL., 1976]	C	4	G		UNSTABLE VARIANCE (<4)	0.74			BITTNER ET AL. (1983); ALLUISI (1969)
ATARI® BASKETBALL	UNKNOWN	C P M	4	I		UNSTABLE (UNSTAB?)	0.58			JONES & KENNEDY (1983) WITH ADAPTATIONS
ATARI® FLAG CAPTURE	UNKNOWN	C P M	1.25	I		30?(30?)	0.58			JONES & KENNEDY (1983) WITH ADAPTATIONS
ATARI® PONG	UNKNOWN	P M	15 APPROX 7 TRIALS	I		UNSTABLE VARIANCE (15?)	0.51			JONES & KENNEDY (1983) WITH ADAPTATIONS
ATARI® RACE CAR	UNKNOWN	P M	1.00	I		90(15)	0.30			KENNEDY ET AL. (1982)
ATARI® SLALOM	UNKNOWN	P M	2.25	I		77(16)	0.36			KENNEDY ET AL. (1982)
ATARI® TOUCH ME	MEMORY SPAN (KENNEDY, ANDREWS CARTER, 1981)	C	7.5	I		150(150)	0.30			KENNEDY, ANDREWS, & CARTER (1981)
AUDITORY DIGIT SPAN (FORWARD)	MEMORY SPAN (MS) (EKSTROM ET AL., 1976)	C	15	G		300(300)	0.38			EKSTROM ET AL. (1976); MCCAFFERTY, BITTNER, & CARTER (1980)
CHOICE REACTION TIME: MOVE-MENT TIME 1	MANUAL DEXTERITY (FLEISHMAN & ELLISON, 1962)	M	5.0 SIMUL. 1-CHOICE	I		UNSTABLE VARIANCE (45)	0.79			KRAUSE & BITTNER (1982);
CHOICE REACTION TIME: MOVE-MENT TIME 2	MANUAL DEXTERITY (FLEISHMAN & ELLISON, 1962)	M	5.0 SIMUL. 2-CHOICE	I		UNSTABLE VARIANCE (45)	0.79			KRAUSE & BITTNER (1982);
CHOICE REACTION TIME: MOVE-MENT TIME 4	MANUAL DEXTERITY (FLEISHMAN & ELLISON, 1962)	M	5.0 SIMUL. 4-CHOICE	I		UNSTABLE VARIANCE (40)	0.86			KRAUSE & BITTNER (1982);
CHOICE REACTION TIME: SLOPE	RATE OF DECISION MAKING (TEICHNER & KREBS, 1974)	P	10.0 1- & 4-CHOICE	I		85(70)	0.41			CARTER & KRAUSE (1983); TEICHNER & KREBS (1974)
COMPLEX COUNTING	SUSTAINED ATTEN-TION (KENNEDY & BRUNS, 1975)	C	15	G		60?(60?)	0.36			KENNEDY & BITTNER (1980); KENNEDY & BRUNS (1975)
FLIGHT SCENARIO: PHANTOMS FIVE®	MIXED (SHAN ON, KRAUSE, & IRONS, 1982)	C P M	10	I		70(70)	0.30			SHANNON ET AL. (1982); GEBELLI (1980)
FREE RECALL	RECALL FROM SHORT TERM MEMORY (UNDERWOOD ET AL., 1977)	C	7 APPROX	G		UNSTABLE VARIANCE (.63)	0.52			HARDESON, KRAUSE, & KENNEDY (1980); FERNANDES & ROSE (1978); UNDERWOOD ET AL. (1977)

*Continued on next page.

TABLE 4. MARGINAL (CONTINUED)*

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE	ADMIN	TOT STAB TIME IN MINUTES (DIFF)	RELIAB	EFFIC	3 MIN	REFERENCES
GRAPHEMIC AND PHONEMIC ANALYSIS: MEAN ERROR TIME	MIXED (ROSE & FERNANDES, 1977)	C	24		G	UNSTABLE (72)	0.23			HARBESON, KENNEDY, ET AL. (1982); BARON (1973); BARON & MCKILLOP (1975); ROSE & FERNANDES (1977)
INTERFERENCE SUSCEPTIBILITY GRAND NUMBER CORRECT (LISTS & SETS)	PROACTIVE INTERFERENCE SUSCEPTIBILITY (UNDERWOOD ET AL., 1977)	C	10		G	80(80)	0.42			KRAUSE & KENNEDY (1980); UNDERWOOD ET AL. (1977)
INTERFERENCE SUSCEPTIBILITY NUMBER CORRECT: LIST 4	ASSOCIATIVE MEMORY (MA) & INTERFERENCE SUSEPTIBILITY (UNDERWOOD ET AL., 1977)	C	10		G	20(20)	0.26			CARTER & KRAUSE (1983); KRAUSE & KENNEDY (1980); UNDERWOOD ET AL. (1977)
LIST DIFFERENTIATION	TEMPORAL SHORT TERM MEMORY (UNDERWOOD ET AL., 1977)	C	6 APPROX		G	UNSTABLE VARIANCE (.48)	0.47			HARBESON ET AL. (1980); FERNANDES & ROSE (1978); UNDERWOOD ET AL. (1977)
MANIKIN TEST: ACCURACY	PICTORIAL CODING ABILITY (EGAN, 1978)	P	7		I	UNSTABLE VARIANCE (49)	0.50			CARTER & WOLDSTAD (IN PRESS); READER ET AL. (1981)
NATURALISTIC VISUAL SEARCH: MEDIAN DETECTION TIME	VISUAL SEARCH	P	14		I	112(112)	0.25			SHANNON ET AL. (IN PRESS)
NAVIGATIONAL PLOTTING: TOTAL CORRECT	MIXED (WIKER, KENNEDY, & PEPPER, 1983)	C P M	9		G	90?(90?)	0.40 APPROX			WIKER ET AL. (1983)
NAVIGATIONAL PLOTTING: TOTAL COMPLETE	MIXED (WIKER ET AL., 1983)	C P M	9		G	90?(90?)	0.40 APPROX			WIKER ET AL. (1983)
SEMANTIC MEMORY RETRIEVAL: PROPERTY, 0-ORDER	PATTERN MATCHING (COLLINS & QUILLIAN, 1969)	P	1.67		G	UNSTABLE VARIANCE (8.37)	0.77			CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)
SEMANTIC MEMORY RETRIEVAL PROPERTY, 1ST-ORDER	PATTERN MATCHING (COLLINS & QUILLIAN, 1969)	P	1.67		G	UNSTABLE VARIANCE (11.69)	0.68			CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)
SEMANTIC MEMORY RETIEVAL: PROPERTY, 2ND-ORDER	RETRIEVAL FROM LTM (COLLINS & QUILLIAN, 1969)	C	1.67			UNSTABLE VARIANCE (5.00)	0.71			CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)

*Continued on next page.

TABLE 4. MARGINAL (CONTINUED)

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE	ADMIN	TOT STAB TIME IN MINUTES (DIFF)	RELIAB	EFF 3 MIN	REFERENCES
SEMANTIC MEMORY RETRIEVAL: SUPERSET, 0-ORDER	RETRIEVAL FROM LTM (COLLINS & QUILLIAN, 1969)	C	1.67	G		UNSTABLE VARIANCE (1.67)	0.77		CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)
SEMANTIC MEMORY RETRIEVAL: SUPERSET, 1ST-ORDER	RETRIEVAL FROM LTM (COLLINS & QUILLIAN, 1969)	C	1.67	G		UNSTABLE VARIANCE (1.67)	0.64		CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)
SEMANTIC MEMORY RETRIEVAL: SUPERSET, 2ND-ORDER	RETRIEVAL FROM LTM (COLLINS & QUILLIAN, 1969)	C	1.67	G		UNSTABLE VARIANCE (18.37)	0.72		CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)
SPOKE DIFFERENCE (C-E) SCORE	MIXED	P M	1.8 APPROX	G		9?(9?)	0.55		BITTNER, LUNDY, ET AL. (1982)
SPOKE EXPERIMENTAL (E) TASK	PERCEPTUAL SPEED (PS) (EKSTROM ET AL., 1976); SPEED, ARM MOVEMENT (FLEISHMAN & ELLISON, 1962)	P M	1.8 APPROX	G		8.3?(8.3?)	0.55		BITTNER, LUNDY, ET AL. (1982); GRAYBIEL, KENNEDY, KNOBLOCK ET AL. (1965)
STROOP: BW-CB DIFFERENCE	COLOR NAMING FACILITY (JENSEN & ROHWER, 1966)	C P	1.0	G		5(1)	0.45		HARBESON, KRAUSE, ET AL. (1982)
STROOP: CW-BW DIFFERENCE	INTERFERENCE PRONENESS (JENSEN & ROHWER, 1966)	C	1.0	G		6(4)	0.47		HARBESON, KRAUSE, ET AL. (1982)
TIME ESTIMATION: CONSTANT ERROR (CE)	PRODUCTION TIME JUDGEMENT (VROON, 1976)	C	15 5 REP. OF 8 INTERVALS	I		180?(180?)	0.88 APPROX		MCCAULEY ET AL. (1980); ZELKIND & SPRUNG (1974)
TRACKING: TWO DIMENSIONAL COMPENSATORY	COMPENSATORY TRACKING, TWO DIMENSIONAL (KENNEDY, BITTNER & JONES, 1981)	P M	1	I		50(50) CEILING DAY 10	0.52		KENNEDY, BITTNER, & JONES (1981); DAMOS ET AL. (1981)
VISUALIZATION	SPATIAL SCANNING (SS) (EKSTROM ET AL., 1976)	P	3	G		UNSTABLE & MEANS VARIANCE (18)	0.66		BITTNER ET AL. (1983); MORAN & MEFFERD (1959)
WONDERLIC PERSONNEL TEST	GENERAL INTELLIGENCE (WONDERLIC, 1978)	C	12	G		48(<12)	0.34		MACKAMAN, BITTNER, HARBESON, KENNEDY, & STONE (1982); WONDERLIC (1978)

TABLE 5. UNACCEPTABLE*

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE	ADMIN	TOT STAB TIME IN MINUTES (DIFF)	RELIAB	EFFIC	3 MIN	REFERENCES
ATARI® BREAKOUT	SLOWLY CHANGING & UNKNOWN	P M	2.00 APPROX	I		VARIANCE (UNSTABLE)	0.41 APPROX			KENNEDY ET AL. (1982)
ATARI® ICE RACE	UNKNOWN	P M	1.00	I		UNSTABLE (UNSTABLE)	0.38 APPROX			JONES & KENNEDY (1983) WITH ADAPTATIONS
AUDITORY DIGIT SPAN (BACKWARDS)	MEMORY SPAN (MS) (EKSTROM ET AL., 1976)	C	15	G		UNSTABLE (UNSTABLE)	0.24 APPROX			EKSTROM ET AL. (1976); MCCAFFERTY ET AL. (1980)
GRAPHEMIC AND PHONEMIC ANALYSIS: SH/HN RATIO	RELATIVE VISUAL/ ACOUSTIC ENCODING (BARON & MCKILLOP, 1975)	C	16	G		UNSTABLE (UNSTABLE)	0.00 APPROX			HARBESON, KENNEDY, ET AL. (1982); BARON (1973); BARON & MCKILLOP (1975); ROSE & FERNANDES (1977)
GRAPHEMIC AND PHONEMIC ANALYSIS: % ERRORS	MIXED (ROSE & FERNANDES, 1977)	C	24	G		192(192)	0.12			HARBESON, KENNEDY, ET AL. (1982); BARON (1973); BARON & MCKILLOP (1975); ROSE & FERNANDES (1977)
INTERFERENCE SUSCEPTIBILITY SLOPE ACROSS LISTS	PROACTIVE INTERFERENCE SUSCEPTIBLITY (UNDERWOOD ET AL., 1977)	C	10	G		UNSTABLE (UNSTABLE)	0.03 APPROX			CARTER & KRAUSE (1983); UNDERWOOD ET AL. (1977)
LETTER CLASSIFICATION: N - P	NAME SEARCH TIME (POSNER & MITCHELL, 1967)	C	24	G		216(216)	0.02			HARBESON, KENNEDY, ET AL. (1982); POSNER & MITCHELL (1967); ROSE & FERNANDES (1977)
LETTER CLASSIFICATION: C - N	CATEGORY SEARCH TIME (POSNER & MITCHELL, 1967)	C	23	G		253(253)	0.10			HARBESON, KENNEDY, ET AL. (1982); POSNER & MITCHELL (1967); ROSE & FERNANDES (1977)
LEXICAL DECISION MAKING: GRAPHEMIC AND PHONEMIC FACILITATION	READING SPEED (MEYER, SCHVANEVELDT, & RUDDY, 1974)	C	3	G		UNSTABLE (UNSTABLE)	0.00 APPROX			KENNEDY & HARBESON (IN PRESS); MEYER ET AL. (1974); ROSE & FERNANDES (1977)
LEXICAL DECISION MAKING: GRAPHEMIC INTERFERENCE	ACOUSTIC OR PHONEMIC ENCODING (MEYER ET AL., 1974)	C	3	G		UNSTABLE (UNSTABLE)	0.00 APPROX			KENNEDY & HARBESON (IN PRESS); MEYER ET AL. (1974); ROSE & FERNANDES (1977)
LEXICAL DECISION MAKING: PHONEMIC SIMILARITY	VISUAL OR GRAPHEMIC ENCODING (MEYER ET AL., 1974)	C	3	G		UNSTABLE (37)	0.27 APPROX			KENNEDY & HARBESON (IN PRESS); MEYER ET AL. (1974); ROSE & FERNANDES (1977)
MAZE TRACING	SPATIAL SCANNING (SS) (EKSTROM ET AL., 1976)	P	2	G		NOT EQUIVALENT	INESTIMABLE			KRAUSE & WOLDSTAD (1983); SHANNON (1982)
NAVIGATIONAL PLOTTING: PERCENT CORRECT	MIXED (WIKER ET AL., 1983)	C P M	9	G		UNSTABLE (UNSTABLE)	INESTIMABLE			WIKER ET AL. (1983)

*Continued on next page.

TABLE 5. UNACCEPTABLE (CONTINUED)

NAME	FACTOR	DOMAIN	ADMIN TIME (MIN)	TYPE	ADMIN	TOT STAB TIME IN MINUTES (DIFF)	RELIAB	EFFIC	3 MIN	REFERENCES
RUNNING RECOGNITION: NUMBER CORRECT	RECOGNITION FROM SHORT TERM MEMORY (UNDERWOOD ET AL., 1977)	C	4 APPROX	G		UNSTABLE (UNSTABLE)	INESTIMABLE			HARBESON ET AL. (1980); FERNANDES & ROSE (1978); UNDERWOOD ET AL. (1977)
SEMANTIC MEMORY RETIEVAL: PROPERTY SLOPE	LTM SCANNING (ROSE & FERNANDES, 1977)	C	3.34	G		UNSTABLE VARIANCES (16.77)	0.00 APPROX			CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)
SEMANTIC MEMORY RETIEVAL: SUPERSET SLOPE	LTM SCANNING (ROSE & FERNANDES, 1977)	C	3.34	G		UNSTABLE VARIANCES (36.7)	0.00 APPROX			CARTER & KRAUSE (1983); COLLINS & QUILLIAN (1969); KENNEDY & HARBESON (IN PRESS); ROSE & FERNANDES (1977)
STERNBERG ITEM RECOGNITION: INTERCEPT	STIMULUS PROCESSING & RESPONSE FORMATION TIME (STERNBERG, 1966, 1975)	P M	12 (4 SET SIZES)	I		UNSTABLE (UNSTABLE)	0.00 APPROX			CARTER ET AL. (1980); CARTER & KRAUSE (1983); STERNBERG (1966, 1975)
STERNBERG ITEM RECOGNITION: SLOPE	SHORT-TERM MEMORY SCAN RATE (STERNBERG, 1966, 1975)	C	12 (4 SET SIZES)	I		UNSTABLE (UNSTABLE)	0.11 APPROX			CARTER ET AL. (1980); CARTER & KRAUSE (1983); STERNBERG (1966, 1975)
TIME ESTIMATION: VARIABLE ERROR	PRODUCTION TIME JUDGEMENT (VROON, 1976)	C	15 (5 REP 8 INTERVALS	I		UNSTABLE (UNSTABLE)	0.35 APPROX			MCCAULEY ET AL. (1980); ZELKIND & SPRUNG (1974)
TRACKING: DUAL CRITICAL-TWO DIMENSIONAL COMPENSATORY	MIXED	P M	1	I		UNSTABLE (UNSTABLE)	0.00 APPROX			KENNEDY ET AL. (1981); DAMOS ET AL. (1981)
VISUAL CONTRAST SENSITIVITY: METHOD OF ADJUSTMENT	CONTRAST SENSITIVITY FUNCTION: 1, 2, 4, 8, 16 cpd (GINSBURG & EVANS, 1982)	P	15 (EACH cpd)	I		UNSTABLE (UNSTABLE)	VARIED			GINSBURG ET AL. (1983); GINSBURG & EVANS (1982)
VISUAL CONTRAST SENSITIVITY: BEKESY METHOD	CONTRAST SENSITIVITY FUNCTION: 1, 2, 4, 8, 16 cpd (GINSBURG & EVANS, 1982)	P	15 (EACH cpd	I		UNSTABLE (UNSTABLE)	VARIED			GINSBURG ET AL. (1983); GINSBURG & EVANS (1982)
VISUAL RESOLUTION ACUITY: ERRORS	VISUAL ACUITY & PERCEPTUAL SPEED	P	1	I		INESTIMABLE	INESTIMABLE			GUIGNARD, BITTNER, EINBENDER, & KENNEDY (1980); GUIGNARD, LANDRUM & REARDON (1976)
VISUAL RESOLUTION ACUITY: TIME	VISUAL ACUITY & PERCEPTUAL SPEED	P	1	I		INESTIMABLE	INESTIMABLE			GUIGNARD ET AL. (1980); GUIGNARD ET AL. (1976)

TABLE 6. MINI-BATTERY FOR ENVIRONMENTAL RESEARCH

NAME	RATIONALE FOR INCLUSION	ADMINISTRATION TIME IN MINUTES	RELIABILITY EFFICIENCY (3 MINUTES)
GRAMMATICAL REASONING	Assesses an analytic cognitive neuropsychological function associated with the left hemisphere.	1.5	0.93
PATTERN COMPARISON:	Assesses an integretive spatial function neuropsychologically associated with the right hemisphere.	2.0	0.93
CODE SUBSTITUTION	This is a mixed associative memory-perceptual speed task which provides for a traditional assessment of these components not otherwise covered by other measures.	2.0	0.84
AIMING	Directly provides for the assessment of environmental effects on fine eye-hand coordination and indirectly provides for separation of such effects from other cognitive measures.	2.0	0.87
SPOKE CONTROL (C) TASK	Directly assesses arm movement speed and indirectly provides for distinction of gross environmental disruptions from disruptions in fine eye-hand coordination and cognition.	<1.0	0.95

END

FILMED

5-85

DTIC